Estimates of station usage April 2022 to March 2023

Frequently Asked Questions

Contents

Background information

The latest in the series of **Estimates of station usage**, for all stations in Great Britain, was published by **Office of Rail and Road** (ORR) on 14 December 2023. This is available on the ORR data portal.

The **Estimates of station usage** statistics consist of estimates of the total numbers of people:

- Travelling from or to the station (**entries and exits**); and
- Changing trains at the station (**interchanges**)

The estimates of entries and exits are further split by ticket type (full price, reduced price and season tickets). There is also a range of station attribute information included too. Further station attribute information can be found in Table 6329 on the data portal.

Alongside these frequently asked questions, a number of documents are available on the data portal, including:

- Statistical release (PDF)
- Passenger entries and exits and interchanges by station (April 2022 to March 2023) – Table 1410 (ods and csv)
- Time series of passenger entries and exits and interchanges by station (April 1997 to March 1998 to April 2022 to March 2023) – Table 1415 (ods)
- Quality and methodology report (PDF)
- Interactive dashboard (Power BI)
- Animated graphics (MP4) and infographics (PDF)

If you have any further enquiries or feedback about the Estimates of station usage, the ORR Information and Analysis Team can be contacted at rail.stats@orr.gov.uk.

What is the reference period for the Estimates of station usage?

The estimates are based on the financial year. Each year runs from 1 April to 31 March. For 1 April 2019 to 31 March 2020, an alternative dataset from 1 March 2019 to 29 February 2020 was also produced.

When will the next dataset be released?

ORR aims to publish the Estimates of station usage as soon as possible after the end of the data reference period. This is usually around eight to nine months after the end of the financial year. It is important to take time to implement the detailed methodology, collect additional data (e.g. from PTEs) and carry out a series of quality assurance checks to ensure the final data are as accurate as possible.

ORR will continue to work with our consultants and stakeholders to shorten the time between the reference period and publication. Publication of Estimates of station usage for the year April 2023 to March 2024 is

expected to be in November 2024. For the most up to date information, please refer to the ORR publication schedule.

Where can historic Estimates of station usage data be found?

Estimates of station usage are available on an annual basis from April 1997 onwards. Annual datasets for the last six years are published on the data portal.

We also publish a collated dataset giving an annual time series of Estimates of station usage from April 1997 to the most recent year. This is also available on the data portal – see Table 1415.

Why are there no Estimates of station usage for April 2003 to March 2004?

As a result of improvements made to the methodology, Estimates of station usage for the year April 2003 to March 2004 were not produced.

Where can data prior to April 1997 be found?

ORR does not hold any estimates of station usage data prior to April 1997.

The National Archives or the National Railway Museum, which holds railway documentation from pre-privatisation, may be able to provide further assistance.

Data scope and methodology

For more detailed information about the methodology used to produce the Estimates of station usage, please refer to the Quality and methodology report.

How is the data for the Estimates of station usage derived?

These statistics on station usage are estimates primarily based on ticket sales, sourced from LENNON, the rail industry's ticketing and revenue system, and local ticketing data. These data sources and the methodology used provide the best approach possible given Great Britain does not have a fully gated rail network or robust count data for every station.

LENNON data feeds into a base matrix which is an input into the MOIRA2.2 rail planning tool. This is supplemented by local ticketing data for Passenger Transport Executive (PTE) areas. These sources are combined, and further adjustments are made to the data to address known issues with the MOIRA2.2 base matrix. These include an allocation of tickets sold to 'London Terminals,' allocation of demand between individual stations in group stations outside of London and a number of cases where adjustments are made to selected stations to account for specific known issues, for example Digby and Sowton. Further information on the data sources adjustments used to estimate usage at individual stations can be found in Table 1410 (column L).

The resulting dataset is used to produce the Origin Destination Matrix (ODM), a comprehensive matrix of passenger flows throughout Great Britain. The ODM is then used to derive estimates for the number of entries and exits at each station in Great Britain. A complete overview of the methodology can be found in the Quality and methodology report.

How can I tell whether the estimate for a specific station is supplemented by local ticketing data or has been adjusted?

The main data table (Table 1410) includes a 'Data source/adjustments' column to flag where the estimate for a specific station has been improved through an adjustment or supplemented by local ticketing data. Further information can be found in the 'Cover_sheet' tab on the table and in the Quality and methodology report.

The data for a particular station is not correct

It is important to emphasise that these data are **estimates** of station usage.

These **estimates** based primarily on tickets sales and the methodology used is the best approach possible given Britain does not have a fully gated rail network or comprehensive and robust count data at every station. However, this data does have weaknesses when utilised for this purpose and, although some of these are catered for in the methodology and we continue to seek improvements to address identified issues, the user should be aware of these acknowledged limitations and bear these in mind when using the data. More information on the limitations of the data and details of the methodology can be found in the Quality and methodology report.

Why does it take so long to produce the Estimates of station usage when ticket sales data is available overnight?

ORR aims to publish the Estimates of station usage as soon as possible after the end of the financial year, which is currently around eight months after. It is important to take time to implement the detailed methodology, collect additional data (e.g. from PTEs) and carry out a series of quality assurance checks to ensure the final data are as accurate as possible.

What changes have been made to the methodology this year?

Improvements to the methodology are made in most years. This is in response to user feedback suggesting that users would prefer the data to be as accurate as possible, rather than maintain absolute consistency in the time series.

As a result, changes to the methodology mean that direct year-on-year comparisons for the stations affected are not valid. The following changes were made to the methodology for the latest year (April 2022 to March 2023):

Introduction of an adjustment to account for split ticketing: Split ticketing is where a passenger completes a single journey using two or more tickets. Total entries and exits at some individual stations will

be significantly over estimated due to some passengers not boarding or alighting at the split point and instead staying on the train. In the April 2021 to March 2022 release we identified the following common ticket split points (sorted alphabetically): Basingstoke, Croydon BR (East Croydon and West Croydon), Didcot Parkway, Doncaster, Gatwick Airport, Leeds, Milton Keynes Central, Peterborough, Sheffield, and York.

For this year we have introduced an adjustment to account for split ticketing in the LENNON data. Rail Delivery Group (RDG) have developed an algorithm to detect split tickets, and we have incorporated this into our data processing. Therefore, in the latest year (April 2022 to March 2023) we only count an entry and exit at the start of the journey and end of the journey, whereas previously an entry and exit would have been recorded at each split point along the way.

Improved estimates of distribution across West Yorkshire Metro area: An updated distribution has been used for allocating journeys using West Yorkshire Metro tickets. The new distribution uses latest year (April 2022 to March 2023) data (replacing the April 2021 to March 2022 data), better reflecting post-pandemic travel patterns.

Improved estimates of concessionary travel across the Greater Manchester area: The methodology for calculating concessionary demand in the Greater Manchester area has been updated. A new data source was available (rail surveys undertaken in August 2023) which covered all ticket types (including concessions on radial Manchester rail routes). Previously survey data from 2017 was used. This data is to estimate the proportion of concessions across all ticket types to account for concessionary travel that would otherwise not be covered by the ticket purchase data.

Improved methodology in how entries and exits are allocated to stations for tickets to London Terminals: An updated distribution has been used for allocating journeys for passengers who have bought non-travelcard tickets with a destination of London Terminals, for example Oxford to London Terminals. The new distribution uses modelled data from the December 2022 CAF to allocate flows to each station based on possible travel routes. Previously these were allocated using results from the 2001 London Area Travel Survey (LATS). While many allocations to London Terminals will be similar to when the LATS took place, some will have changed significantly (e.g. due to the impacts of Thameslink). This represents a significant methodological improvement.

For more information about these changes and the impact of improving the methodology, please refer to the Quality and methodology report.

How is usage of group stations estimated?

Group stations are a small number of stations that are treated as a group for an origin or destination, rather than individual stations. For example, where a ticket identifies the origin or destination as Manchester BR, passengers could use Manchester Piccadilly, Manchester Victoria, Manchester Oxford Road, Salford Central or Deansgate.

Current industry data does not distinguish between the component stations and therefore a split of entries and exits between these stations has been estimated during the production of Estimates of station usage.

Why are journeys for certain operators excluded from the data?

Sales for Eurostar services are not included in the rail industry ticketing systems. Therefore, travel using tickets from this operator is not included in Estimates of station usage. Consequently, our estimates of usage at St. Pancras will not be a true reflection of the total usage at this stations.

What ticket types are included in the data and which category do these tickets belong in?

- Full – all walk-up **undiscounted** single or return tickets, whether or not issued with a status discount (child, railcard, etc).
- Reduced – all walk-up **discounted** single or return tickets, whether or not issued with a status discount (child, railcard, etc). All advance-purchase tickets are also included in the category.
- Seasons – all multi-use tickets

How are sales of different ticket types converted into entries and exits?

Ticket transactions are converted into an estimate of the number of journeys made by applying a series of ticket type journey factors.

Single and return tickets unambiguously translate into one and two journeys respectively. The number of journeys made using season tickets is estimated using a historic estimate of the monthly use of a season ticket. Ticket periods of other lengths are converted into a number of journeys using a proportion of the monthly journey factor, e.g. monthly seasons have a journey factor of 45 and annual seasons a journey factor of 480.

Does the data include people who use the station but do not travel?

This is defined as ‘station footfall’ and includes those individuals entering the station without the intention of travelling on the rail network. This includes use of shopping or restaurant facilities.

As Estimates of station usage are based on tickets sales, the data is representative of those individuals that are intending to travel on the rail network. Therefore, station footfall is not included in the data.

Network Rail publishes information on station footfall at Network Rail managed stations only.

Does the data include those people who travel on the train without purchasing a ticket?

This is defined as ‘ticketless travel’.

As Estimates of station usage are based on ticket sales, journeys associated with ticketless travel are not included in the data. This is more likely to be an issue on some flows and where ticketless travel is significant as more stations have been gated over time and train operators focus on revenue protection

activities, this is likely to be less of an issue than in the past. It should be noted that levels of ticketless travel may have changed during the pandemic and those changes may vary substantially by station.

There is a strong argument that it is not appropriate to include ticketless travel in the dataset as its purpose is to record genuine journeys on the rail network. The inclusion of ticketless travel could distort business cases for new investment where these are reliant on Estimates of station usage data.

It is worth noting that ticketless travel also includes an element of individuals who are legitimately travelling for free, such as the British Transport Police or some rail industry employees.

Can a further breakdown of the Estimates of station usage be produced?

These estimates are annual (financial year) only. Using the current methodology and, in particular, the use of the MOIRA2.2 base matrix, it is not possible to produce a further temporal breakdown, such as by week or time of day. In addition, the source data from LENNON only provides details on when a ticket was purchased rather than when the passenger actually travelled. Therefore, journeys cannot be assigned to specific days or times.

It is not possible to provide a breakdown of the data by traveller type, such as business, leisure or commuter, as there is no way of identifying this information within the data. Some information on the reasons for travel by rail can be found in the National Travel Survey, which is published by the Department for Transport.

Which geographical boundary definitions have been used?

One geographical definition is included in the main table for each station: Region (boundary definitions as of April 2022).

Another table published on our data portal (Table 6329) contains geographic and other attribute information for each mainline station, including location coordinates (Easting and Northing), county, constituency and station facility owner.

More information on these boundary definitions, as well as the boundary files, can be found on the Office for National Statistics Open Geography Portal.

Is it possible to provide an alternative geographic breakdown?

ORR is not able to provide bespoke geographical breakdowns of the Estimates of station usage statistics.

However, as both the local authority and the coordinates (Easting and Northing) of each station are available, it is possible to create an alternative geographical definition, using either a lookup file or a Geographical Information System (GIS), such as QGIS or ESRI ArcGIS.

More information on geographical definitions in the United Kingdom, including look up files and boundary files to download, is available on the Office for National Statistics Open Geography Portal.

Why doesn't the dataset include information about the number of platforms or services that call at each station?

We are unable to accurately source this information. We welcome any views on whether this would be useful or if there is other information at a station level that would add value to our station attributes table (Table 6329).

Why are there some stations with no data?

In general, a station is recorded in the Estimates of station usage statistics if there is a record of ticket sales with the station as an origin or destination. However, it is known that there are some national rail stations for which sales are not recorded in the LENNON ticketing system, such as stations only used on a very limited number of days, e.g. Bishops Lydeard or Corfe Castle.

To be consistent with the total station count in the Rail Infrastructure and Assets statistical release as at 31 March 2023, these stations have been included in the station list in Table 1410, but have no data shown. Stations with no usage estimates available for any year since April 1997 are not included in the time series dataset (Table 1415).

Why are there some stations missing when the Estimates of station usage has data for all stations?

The Estimates of station usage dataset includes all stations served by national rail services as of 31 March 2023. Data for any stations opened after 31 March 2023 will be included in the next publication.

If you require information for a non-national rail station, please contact the relevant station operator or owner, who may be able to provide further assistance.

Using the Estimates of station usage dataset

Where are the columns in the dataset defined?

A full description of the column headings can be found in the 'Cover_sheet' tab of the Estimates of station usage data tables.

For more information about these fields, please refer to the Quality and methodology report.

Is it possible to search for a specific station?

Yes. Users can add filters to the Estimates of station usage data tables that can be used to search for a specific station. The interactive dashboard on the estimate of station usage page of the data portal can also be easily filtered.

Related Statistics

Where can I find information of journeys between pairs of stations

The Origin and Destination Matrix (ODM) which contains the number of journeys between every combination of mainline stations in Great Britain during April 2022 to March 2023 will be published on the Rail Data Marketplace in February 2024, following the release of our Regional Rail Usage publication. The ODM covering April 2021 to March 2022 is available to download.

Where can information on station footfall be found?

Network Rail publishes information on station footfall at Network Rail managed stations only.

Where can information about station car park usage be found?

For information about station car park usage, please contact the relevant train operator who manages the station.

Where can I find information on the number of rail journeys made on each train operator?

ORR publishes the number of journeys by train operator every quarter in the Passenger Rail Usage statistical release.

Where can I find information on train crowding?

DfT publishes annually information on rail passenger numbers and crowding during a typical autumn weekday for fourteen major cities across England and Wales, as well as central London stations, in its Rail passenger numbers and crowding statistical release.